Stochastic Processes and their Applications 27 (1988) 141-149
North-Holland

ON THE SIZE OF A RUMOUR

Ray WATSON
Department of Statistics, University of Melbourne, Parkville, VIC 3052, Australia

Received 2 February 1987
Revised 20 July 1987

Two rumour models are considered, the first due to Daley and Kendall (1965) and the second due to Maki and Thompson (1973). The size of a rumour, defined as the number of individuals in the population eventually hearing the rumour, is investigated using both deterministic and stochastic approximations, and some asymptotic distributional results for the size distribution for the two models are obtained and compared.

rumour * deterministic * stochastic * martingale * approximation

1. Introduction

The spreading of information is in many ways similar to the spreading of infection. Thus it is not surprising that epidemic models—particulalrly the standard epidemic model—have been used to describe the spread of information in the form of rumours. Of course, there are also differences—most obviously in the mechanism of removal, i.e. the event that a spreader ceases to spread the rumour. We consider first the rumour model due to Daley and Kendall (1965) and subsequently investigated by Barbour (1972). This model assumes that a spreader continues to spread the information until meeting someone who has heard it: either another spreader, in which case both cease spreading, or a removed case.

This assumption removes the need for a separate removal rate: both infection and removal rates depend only on the rate at which meetings between individuals take place and on the rules for spread of information.

As one might expect, the behaviour of the model is similar to that of the standard epidemic model, and yet different. There is no threshold since the initial relative removal rate is effectively about half the population size; thus the model is necessarily supercritical. In Section 2 we consider the behaviour of this model in some detail, and use a martingale central limit theorem to derive an approximation for the size distribution. Also we verify some conjectures made by Daley and Kendall (1965) concerning the size distribution. Maki and Thompson (1973) proposed an alternative rumour model, which was more recently considered by Frauenthal (1980) and

0304-4149/88/$3.50

Sudbury (1985). This model is actually a simplified version of the Daley-Kendall model in which the double removal transition is not allowed. Instead, it is assumed that if two spreaders meet, only one is removed. The deterministic approximation to this model is exactly the same as for the Daley-Kendall model. However, the modified removal assumption simplifies the martingale analysis of the stochastic model considerably. This model is considered in Section 3 where results analogous to those for the Daley-Kendall model are obtained.

2. The Daley-Kendall model

We adhere to the notation and terminology used for epidemic models. Thus we assume that at time t there are $X(t)$ susceptibles (ignorant of the rumour), $Y(t)$ infectives (spreaders), and $Z(t)$ removed cases (non-ignorant non-spreaders, or stiflers). Initially $X(0)=n$, $Y(0)=a$ and $Z(0)=0$. The population is assumed closed, so that $X(t)+Y(t)+Z(t)=n+a$.

The process $\{(X(t), Y(t), Z(t)), t \geqslant 0\}$ has transition probabilities given by

transition in $(t, t+\mathrm{d}t)$		transition probability
$X(t+\mathrm{d}t)=X(t)-1,$	$Y(t+\mathrm{d}t)=Y(t)+1$	$n^{-1}\beta X(t)Y(t)\,\mathrm{d}t$
$Y(t+\mathrm{d}t)=Y(t)-2,$	$Z(t+\mathrm{d}t)=Z(t)+2$	$\frac{1}{2}n^{-1}\beta Y(t)[Y(t)-1]\,\mathrm{d}t$
$Y(t+\mathrm{d}t)=Y(t)-1,$	$Z(t+\mathrm{d}t)=Z(t)+1$	$n^{-1}\beta Y(t)[n+a-X(t)-Y(t)]\,\mathrm{d}t$

with initial conditions $X(0)=n$, $Y(0)=a$, $Z(0)=0$.

We choose rate constant $n^{-1}\beta$ on the assumption that each individual contacts other individuals at a constant rate per unit time, irrespective of population size. This facilitates comparison of time scales of populations of different sizes.

The deterministic approximation for this process is such that

$$\frac{\mathrm{d}x}{\mathrm{d}t}=-n^{-1}\beta xy,$$

$$\frac{\mathrm{d}y}{\mathrm{d}t}=n^{-1}\beta(2x-n-a+1)y,$$

$$\frac{\mathrm{d}z}{\mathrm{d}t}=n^{-1}\beta(n+a-1-x)y.$$

From the first and third of these equations we obtain

$$\frac{\mathrm{d}z}{\mathrm{d}x}=-\frac{n+a-1}{x}+1$$

so that the function

$$k(t)=n-x(t)+z(t)-(n+a-1)\ln[n/x(t)] \tag{1}$$

is a constant of the motion, with value 0.

Now if we let $n\to\infty$ and $a\sim n\phi$ where $\phi>0$, then we have, following Kurtz (1970) and Barbour (1972),

$$(n^{-1}X(t), n^{-1}Y(t), n^{-1}Z(t))\xrightarrow{\mathrm{P}}(u(t), v(t), w(t)) \tag{2}$$

where u, v and w are such that

$$\frac{\mathrm{d}u}{\mathrm{d}t}=-\beta uv,$$

$$\frac{\mathrm{d}v}{\mathrm{d}t}=\beta(2u-1-\phi)v,$$

$$\frac{\mathrm{d}w}{\mathrm{d}t}=\beta(1+\phi-u)v,$$

with $u(0)=1$, $v(0)=\phi$, $w(0)=0$. While these equations are similar to the corresponding equations for the standard epidemic model, there are significant differences.

The deterministic approximation for the extent, e, of the outbreak (i.e. the proportion of the population who eventually hear the rumour) is obtained by putting $x(\infty)=n(1-e)$, $y(\infty)=0$ and $z(\infty)=ne+a$ in (1). This gives

$$\ln(1-e)+\frac{2n}{n+a-1}\left(e+\frac{a}{2n}\right)=0. \tag{3}$$

The severity, e, of the standard epidemic model with threshold parameter θ and initial proportion of infectives δ is given by

$$\ln(1-e)+\theta(e+\delta)=0.$$

If $\delta>0$ then this equation has a positive root which we denote by $\varepsilon(\theta,\delta)$. Note that if $\theta<1$ then ε is near zero (minor outbreak) while if $\theta>1$ then ε is away from zero (major outbreak). Thus we see that the deterministic approximation for the extent of an outbreak of a rumour is nearly the same as the deterministic approximation for the severity of an outbreak for the standard epidemic model with threshold parameter $\theta\approx 2$ and inital proportion of defectives $\delta\approx\frac{1}{2}\phi$:

$$e=\varepsilon\left(\frac{2n}{n+a-1},\frac{a}{2n}\right).$$

A detailed table of values of the deterministic approximation for the extent is given in Table 1. It is seen that the extent is quite stable at around 0.8 for a range of values of n and a. It is also seen, as remarked upon by Daley and Kendall (1965), that the extent actually decreases as the number of initial spreaders increases. This

Table 1
Deterministic approximation for the proportion of initially susceptible individuals who eventually hear the rumour

e	$a=1$	$a=2$	$a=5$	$a=10$	$a=20$	$a=50$	$a=100$
$n=100$	0.8002	0.7982	0.7923	0.7832	0.7672	0.7326	0.7009
$n=200$	0.7985	0.7975	0.7945	0.7897	0.7807	0.7579	0.7307
$n=500$	0.7975	0.7971	0.7959	0.7939	0.7900	0.7792	0.7634
$n=1000$	0.7972	0.7970	0.7963	0.7953	0.7934	0.7876	0.7786
$n=5\,000$	0.7969	0.7968	0.7967	0.7965	0.7961	0.7949	0.7929
$n=10\,000$	0.7968	0.7968	0.7968	0.7967	0.7965	0.7965	0.7949

is a little misleading however, since the proportion of the total population who eventually know the rumour, given by

$$e^* = \frac{n\varepsilon\left(\dfrac{2n}{n+a-1}, \dfrac{a}{2n}\right) + a}{n+a}$$

does increase with a. Further it is even more stable than e, as indicated by Table 2.

In the case of a supercritical standard epidemic process there remains a possiblity of a minor outbreak. The probability of a minor outbreak is approximated by using a linear birth and death process to approximate the initial behaviour of the epidemic process. However this technique is inappropriate for the rumour model.

To determine the minor outbreak distribution, it is necessary to enumerate possible paths in the random walk. This indicates that the probability of a minor outbreak is $O(n^{-a})$ for $a \ll n$, as shown by Daley and Kendall (1965). Provided n is not small this probability will be negligible provided $a > 1$. In all that follows, the possibility of a minor outbreak is ignored. The results can be treated as conditional on a major outbreak and, if required, to more accurately represent the size distribution, a minor outbreak distribution may be mixed with the major outbreak results as in Watson (1980).

Table 2
Deterministic approximation for the proportion of the total population who eventually hear the rumour

e^*	$a=1$	$a=2$	$a=5$	$a=10$	$a=20$	$a=50$	$a=100$
$n=100$	0.8022	0.8022	0.8022	0.8029	0.8060	0.8217	0.8504
$n=200$	0.7995	0·7995	0.7995	0.7997	0.8006	0.8063	0.8205
$n=500$	0.7979	0.7979	0.7979	0.7979	0.7981	0.7992	0.8028
$n=1000$	0.7974	0.7974	0.7974	0.7974	0.7974	0.7977	0.7978
$n=2000$	0.7971	0.7971	0.7971	0.7971	0.7971	0.7972	0.7975
$n=5\,000$	0.7976	0.7976	0.7976	0.7976	0.7976	0.7976	0.7970
$n=10\,000$	0.7969	0.7969	0.7969	0.7969	0.7969	0.7969	0.7969

To approximate the behaviour of the model in the (asymptotically certain) event of a major outbreak, we define the process

$$K(t)=n-X(t)+Z(t)-(n+a-1)[H_1(n)-H_1(X(t))] \tag{4}$$

where $H_1(\xi)=\sum_{j=1}^{\xi} j^{-1}$, which is clearly based on the deterministic constant of the motion (1).

The process $\{K(t),\ t\geqslant 0\}$ has transition probabilities given by

transition in $(t, t+\mathrm{d}t)$	transition probability
$K(t+\mathrm{d}t)=K(t)+1-\dfrac{n+a-1}{X(t)}$	$n^{-1}\beta X(t)Y(t)\,\mathrm{d}t$
$K(t+\mathrm{d}t)=K(t)+2$	$\frac{1}{2}n^{-1}\beta Y(t)[Y(t)-1]\,\mathrm{d}t$
$K(t+\mathrm{d}t)=K(t)+1$	$n^{-1}\beta Y(t)[n+a-X(t)-Y(t)]\,\mathrm{d}t$

with initial condition $K(0)=0$. Thus we see that $\{K(t),\ t\geqslant 0\}$ is a martingale. A central limit theorem as $n\to\infty$ is readily proved following the method used by Watson (1981). Here there is no need to consider the problem of early extinction, and all that is required is to verify the asymptotic negligibility of the squared martingale increments. This reduces to showing that $E[n(X_n(T)+1)^{-2}]\to 0$ as $n\to\infty$, where $X_n(T)$ denotes the final number of susceptibles at the end of a rumour for which the intial number of susceptibles is n. This follows, essentially because $X_n(T)$ remains of the order of n: the details follow Watson (1981).

Thus we obtain

$$\frac{K}{U}\xrightarrow{\mathrm{d}} N,\quad \frac{K}{\sigma}\xrightarrow{\mathrm{d}} N \qquad \text{as } n\to\infty \tag{5}$$

where N denotes the standard normal distribution, with

$$U^2=\sum(\Delta K)^2$$

i.e. the sum of the squared martingale increments. Thus

$$U^2=\sum_{j=X+1}^{n}\left(1-\frac{n+a-1}{j}\right)^2+4R+S$$

where $R(t)$ denotes the number of transitions $K\to K+2$ (corresponding to the meeting of two spreaders) up to time t, and $S(t)$ denotes the number of transitions $K\to K+1$ (corresponding to the meeting of a spreader and a stifler) up to time t. Now, since $Z(t)=S(t)+2R(t)$, we obtain

$$\begin{aligned}U^2&=n-X-2(n+a+1)[H_1(n)-H_1(X)]\\&\quad+(n+a-1)^2[H_2(n)-H_2(X)]+Z+2R\end{aligned}$$

where $H_2(\xi)=\sum_{j=1}^{\xi} j^{-2}$. Also, $\sigma^2=E(U^2)=\mathrm{var}(K)$.

We note that it is convenient to replace $H_1(n)-H_1(x)$ and $H_2(n)-H_2(x)$ by asymptotically equivalent forms:

$$H_1(n)-H_1(x)\sim\ln\left(\frac{n}{x}\right),\qquad H_2(n)-H_2(x)\sim\frac{1}{x}-\frac{1}{n}.$$

The result in the form (5) is of little use however, since the process $\{R(t), t\geq 0\}$ is usually unobservable. We therefore seek an alternative estimator of the variance: if S^2 denotes an estimator of σ^2 such that $S/\sigma\xrightarrow{\text{p}}1$ as $n\to\infty$, then we have

$$\frac{K}{S}\xrightarrow{\text{d}}N\quad\text{as } n\to\infty.$$

To derive such an estimator of the variance, we consider the approximating process $\{K^{\#}(t), t>0\}$ which has deterministic transition probabilities given by

transition in $(t, t+\mathrm{d}t)$	transition probability
$K^{\#}(t+\mathrm{d}t)=K^{\#}(t)+1-\dfrac{n+a-1}{x(t)}$	$n^{-1}\beta x(t)y(t)\,\mathrm{d}t$
$K^{\#}(t+dt)=K^{\#}(t)+2$	$\frac{1}{2}n^{-1}\beta y(t)[y(t)-1]\,\mathrm{d}t$
$K^{\#}(t+\mathrm{d}t)=K^{\#}(t)+1$	$n^{-1}\beta y(t)[n+a-x(t)-y(t)]\,\mathrm{d}t$

with initial condition $K^{\#}(0)=0$.

If $v^{\#}(t)=\operatorname{var}[K^{\#}(t)]$, then $v^{\#}$ is such that

$$\frac{\mathrm{d}v^{\#}}{\mathrm{d}t}=\left(1-\frac{n+a-1}{x}\right)^2\frac{\beta}{n}xy+\frac{2\beta}{n}y(y-1)+\frac{\beta}{n}y(n+a-x-y)$$

Hence,

$$\frac{\mathrm{d}v^{\#}}{\mathrm{d}x}=-\frac{(n+a-1)^2}{x^2}-\frac{n}{x}+2+\frac{n+a-1}{x}\ln\frac{n}{x},\tag{6}$$

and thus

$$v^{\#}=(n+a-1)^2\left(\frac{1}{x}-\frac{1}{n}\right)+n\ln\left(\frac{n}{x}\right)-2(n-x)-\tfrac{1}{2}(n+a-1)\ln^2\left(\frac{n}{x}\right).$$

Thus, an estimator of the variance $\sigma^2=\operatorname{var}[K(t)]$ is given by

$$S^2=(n+a-1)^2\left(\frac{1}{X(t)}-\frac{1}{n}\right)+n\ln\left(\frac{n}{X(t)}\right)-2(n-X(t))$$
$$-\tfrac{1}{2}(n+a-1)\ln^2\left(\frac{n}{X(t)}\right)$$

We note that by rewriting (6) as

$$\frac{\mathrm{d}v^{\#}}{\mathrm{d}x}=-1+\frac{2(n+a-1)}{x}-\frac{(n+a-1)^2}{x^2}-\frac{n+a-x-1}{x}-\frac{y-1}{x}$$

we obtain

$$v^{\#} = n - x - 2(n+a-1)\ln\left(\frac{n}{x}\right) + (n+a-1)^2\left(\frac{1}{x} - \frac{1}{n}\right) + z + 2r$$

which corresponds to the expression for the variance estimator U^2. From the convergence in probability of the process to the deterministic approximations as indicated by (2), we deduce that

$$\frac{U^2}{v^{\#}} \xrightarrow{\mathrm{p}} 1 \quad \text{and} \quad \frac{S^2}{v^{\#}} \xrightarrow{\mathrm{p}} 1$$

And since $U/\sigma \xrightarrow{\mathrm{p}} 1$, it follows that $S/\sigma \xrightarrow{\mathrm{p}} 1$ as required. Thus we have the result that, as $n \to \infty$

$$\frac{K(t)}{S(t)} \xrightarrow{\mathrm{d}} N,$$

$$\frac{n - X(t) + Z(t) - (n+a-1)\ln\left(\dfrac{n}{X(t)}\right)}{\sqrt{(n+a-1)^2\left(\dfrac{1}{X(t)} - \dfrac{1}{n}\right) + n\ln\dfrac{n}{X(t)} - 2[n - X(t)] - \frac{1}{2}(n+a-1)\ln^2\dfrac{n}{X(t)}}} \xrightarrow{\mathrm{d}} N.$$

At the end of the outbreak, when $X(T) = n(1-E)$, $Y(T) = 0$ and $Z(T) = n(E+\phi)$, where E denotes the proportion of initially ignorant individuals who learnt the rumour, the above result gives

$$\frac{\sqrt{n}[(1+\phi)\ln(1-E) + 2E + \phi]}{\sqrt{(1+\phi)^2\left(\dfrac{1}{1-E} - 1\right) - \ln(1-E) - 2E - \frac{1}{2}(1+\phi)\ln^2(1-E)}} \xrightarrow{\mathrm{d}} N.$$

If the initial number of infectives is held constant as $n \to \infty$, so that $a/n \to 0$, we have the simpler result

$$\sqrt{n}\psi(E) = \frac{\sqrt{n}[\ln(1-E) + 2E]}{\sqrt{\left(\dfrac{1}{1-E} - 1\right) - \ln(1-E) - 2E - \frac{1}{2}\ln^2(E)}} \xrightarrow{\mathrm{d}} N \tag{7}$$

Now, if $T_n = \xi_n/\sqrt{n}$, where $\xi_n \xrightarrow{\mathrm{d}} N$, then we obtain $g(T_n) \overset{\mathrm{d}}{\sim} g(0) + \xi_n g'(0)/\sqrt{n}$, by expanding g about 0. It follows that

$$\sqrt{n}\psi'(e)(E-e) \xrightarrow{\mathrm{d}} N$$

where e is such that $\psi(e) = 0$. Moreover, we observe that the function ψ is nearly linear near e, so that for moderate values of n, the distribution of E will also be quite close to normal. Further, $\psi'(e) = 1.794108$, so that $\mathrm{var}(E) \sim 0.310673\, n^{-1}$, which agrees quite closely with the result obtained by Daley and Kendall (1965).

However, the result (7) gives the better approximation to the distribution of E, as the distribution of $\psi(E)$ is less skew than that of E.

3. The Maki–Thompson model

The rumour model proposed by Maki and Thompson (1973) is equivalent to the process $\{(X, t), Y(t), Z(t)), t \geq 0\}$, with transition probabilities given by

transition in $(t, t+dt)$		transition probability
$X(t+\mathrm{d}t)=X(t)-1,$	$Y(t+\mathrm{d}t)=Y(t)+1$	$n^{-1}\beta X(t)Y(t)\,\mathrm{d}t$
$Y(t+\mathrm{d}t)=Y(t)-1,$	$Z(t+\mathrm{d}t)=Z(t)+1$	$n^{-1}\beta Y(t)[n+a-X(t)-1]\,\mathrm{d}t$

with initial conditions $X(0)=n$, $Y(0)=a$, $Z(0)=0$.

This model may be obtained from the Daley-Kendall model by replacing the transition $Y(t+\mathrm{d}t)=Y(t)-2$, $Z(t+\mathrm{d}t)=Z(t)+2$ with rate $\frac{1}{2}n^{-1}\beta Y(t)[Y(t)-1]$ by the transition $Y(t+\mathrm{d}t)=Y(t)-1$, $Z(t+\mathrm{d}t)=Z(t)+1$ with rate $n^{-1}\beta Y(t)$-$[Y(t)-1]$, i.e. the magnitude of the transition is halved and the rate is doubled. This means that the deterministic approximation for this process is identical to the deterministic approximation for the Daley-Kendall process and so all the results of the previous section concerning the deterministic approximation apply equally to the Maki-Thompson model.

Further, the process

$$K(t)=n-X(t)+Z(t)-(n+a-1)[H_1(n)-H_1(X(t))]$$

is a martingale for the MK process as well, and

$$\frac{K}{U}\xrightarrow{\mathrm{d}}N,\ \frac{K}{\sigma}\xrightarrow{\mathrm{d}}N \quad \text{as } n\to\infty$$

where

$$U^2=\sum(\Delta K)^2=n-X+Z-2(n+a-1)\sum_{X+1}^{n}\frac{1}{j}+(n+a-1)^2\sum_{X+1}^{n}\frac{1}{j^2},$$

and

$$\sigma^2=E(U^2)=\operatorname{var}(K).$$

In this case U is directly observable. However a simpler alternative variance estimate is given by

$$S^2=(n+a-1)^2\left(\frac{1}{X(t)}-\frac{1}{n}\right)-(n+a-1)\ln\frac{n}{X(t)}$$

which is obtained by replacing K by zero in the expression for U^2, since $E(K)=0$; and replacing H_1 and H_2 by asymptotic equivalents.

Thus, for the MK model we have the result that, as $n \to \infty$

$$\frac{n - X(t) + Z(t) - (n+a-1)\ln \dfrac{n}{X(t)}}{\sqrt{(n+a-1)^2\left(\dfrac{1}{X(t)} - \dfrac{1}{n}\right) - (n+a-1)\ln \dfrac{n}{X(t)}}} \xrightarrow{\mathrm{d}} N$$

and, at the end of the outbreak

$$\frac{\sqrt{n}[(1+\phi)\ln(1-E) + 2E + \phi]}{\sqrt{(1+\phi)^2\left(\dfrac{1}{1-E} - 1\right) + (1+\phi)\ln(1-E)}} \xrightarrow{\mathrm{d}} N$$

If $\phi = 0$, then we obtain

$$\sqrt{n}\zeta(E) = \sqrt{n}\frac{\ln(1-E) + 2E}{\sqrt{\dfrac{1}{1-E} - 1 + \ln(1-E)}} \xrightarrow{\mathrm{d}} N$$

from which it follows that

$$\sqrt{n}\zeta'(e)(E-e) \xrightarrow{\mathrm{d}} N$$

where e is such that $\ln(1-e)+2e=0$, i.e. $e \approx 0.7968$. Further, $\zeta'(e) \approx 1.9148$, so that $\mathrm{var}(E) \sim 0.2728\, n^{-1}$.

Thus the size distribution is asymptotically normal with mean $e \approx 0.7968$, as for the DK model but with a slightly smaller variance: $0.2728\, n^{-1}$ compared to $0.3107\, n^{-1}$. This accords with the change to smaller more frequent jumps in the removal process.

References

A.D. Barbour, The principle of diffusion of arbitrary constants, J. Appl. Prob. 9 (1972) 519-541.

D.J. Daley and D.G. Kendall, Stochastic rumours, J. Inst. Maths. Applics. 1 (1965) 42-55.

J.C. Frauenthal, Mathematical Modelling in Epidemiology (Springer-Verlag, Berlin, 1980).

T.C. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, J. Appl. Prob. 7 (1970) 49-58.

D.P. Maki and M. Thompson, Mathematical Models and Applications (Prentice-Hall, Englewood Cliffs, NJ, 1973).

A. Sudbury, The proportion of the population never hearing a rumour, J. Appl. Prob. 22 (1985) 443-446.

R.K. Watson, On the size distribution for some epidemic models, J. Appl. Prob. 17 (1980) 912-921.

R.K. Watson, An application of a martingale central limit theorem to the standard epidemic model, Stoch. Proc. Applicns. 11 (1981) 79-89.